SDC

SP-1262

Is Relevance an Adequate Criterion in Retrieval System Evaluation?

Lauren B. Doyle

1 July 1963

(SP Series)

SP-1262

Is Relevance an Adequate Criterion in Retrieval System Evaluation?

Lauren B. Doyle

July 1, 1963

SYSTEM DEVELOPMENT CORPORATION, SANTA MONICA, CALIFORNIA

(page 2 blank)

ABSTRACT

It is argued that the use of "relevance to a search request" as a criterion of what a system retrieves is, in effect, a suboptimization on the machine side of the man-machine interface, and that the searcher needs an efficient exploratory system rather than a request-implementing system.

IS RELEVANCE AN ADEQUATE CRITERION IN RETRIEVAL SYSTEM EVALUATION?

Many advances in science have come as a result of questioning concepts and assumptions which have previously been taken for granted. Copernicus, Darwin, and Heisenberg are noted as much for the ideas they swept away as for the new ideas they brought forth. Today we are quite comfortable with the thought that the earth is not the center of the universe; but this notion died hard in the century of Copernicus. And just as the 14th century philosophers constructed epicycles to maintain the integrity of the earth-centered planetary system, so, in general, will people perform near-unreasonable mental gyrations to defend a conceptual habit which new knowledge and events are making untenable.

The concept of "relevance" has gained in importance in recent times along with the trend toward tighter evaluation of retrieval systems. In evaluation one has to compare system performance to some "ideal" or other kind of standard, and in the case of document retrieval the ideal of performance has sometimes been put: "To retrieve all and only those documents the searcher would regard as relevant to his need if he could personally inspect every document in the library." To duplicate this ideal in the practice of evaluation, it has been customary to assemble judges to inspect documents and concur among themselves as to what documents the system should retrieve in response to a given request.

But this procedure contains an enormous hidden flaw, implicit in the foregoing phrase "in response to a given request," which may cause many current system evaluations to be looked upon--in the course of history--as comparable to the epicycles of the 14th century. Other casualties may be relevance, scales of relevance, and most definitely relevance numbers. The flaw is that there may be a great difference between relevance to a given request statement and relevance to a person's real information need. It is a hidden flaw because in a real information search it is the request statement--the outward expression of the information need--rather than the need itself which comes to the surface.

If the search request statement is usually only an approximate statement of the searcher's need, then the subset of documents pronounced by the judges as relevant to the request would usually be different from that which the searcher himself would choose "if he could personally inspect every document." Taylor (1) is one of several people who have recently pointed out the distinction between relevance to a need and relevance to a request. His portrayal of what the situation might be like is interesting: "Several inquirers may, for example, ask the same question--or what is assumed to be the same question--and each receives, as answer, the same set of messages. However, on scanning, each person picks out a different subset of the total package...as relevant to his question. Although their verbally stated questions were the same, it is obvious that each inquirer must have had a different need. Yet we prescribe the same medicine for each."

If Taylor's imaginary situation were materialized--as some sort of experiment--we would probably find the subsets greatly overlapping; surely there would be a few documents that all the inquirers would agree were relevant. Nevertheless, we still ought to be deeply concerned about the non-overlapping parts of such subsets, because they would reflect the general inability of searchers to ask the right questions of the system.

There exists a lurking feeling among retrieval thinkers, which is hard to dispel by argument, that a searcher's initial try at constructing a request is seldom more than a crude approximation of whatever mechanism will give him the document subset that optimally fulfills his need. Taylor, along with Bar-Hillel (2) and Cheydleur (3), believes that feedback from the information store (or its representation) must be provided by the retrieval system, so that the searcher can redefine his need in a series of iterations (this very process frequently occurs whenever a searcher goes to a human expert in pursuit of information). Stiles (4) revealed that he too is experimenting with what might be called "man-system interaction." This is particularly interesting because Stiles was one of the first to try out associative methods of machine searching, wherein a computer yields not only references which directly satisfy a searcher's request but also references whose index tags are strongly associated (statistically) with the tags making up the request. In other words, even in Stiles' system, which was designed to use statistics in getting around the problem of search-request inadequacy, there is still felt to be a need for feedback.

In visualizing systems of this kind we can feel the usefulness of the concept of relevance slipping through our fingers. We now become aware that the "most relevant subset" is not only an individual matter for the searcher, dependent on the time and circumstances of his searching foray, but also that the feedback he gets is quite capable of changing his idea of what he wants as well as changing his way of expression. An "information need" is thus revealed to be a dynamic entity, whose times of greatest dynamism and change may come in the very process of interacting with a retrieval system.* How does one make use of the concept of relevance in a situation like this?

No, I am afraid that once more, as in the days when the earth was regarded as the center of the universe, we are succumbing to egocentricity when we view a man as commanding a machine to do his bidding in searching the literature. (This is the essence of the "request.") The question is not only whether a man knows how to command,but whether command is at all an appropriate kind of interaction between man and retrieval system. When we read a book, do we command it? No, we either follow it or explore it. Exploratory capabilility, as it turns out, is provided by traditional libraries, but not by some of our modern machine literature searching schemes.

* People who consult permuted title indexes, where feedback in effect is frozen into the structure, are often led astray and find themselves asking: "Now, let's see, what was it I started out to look for?"

Also, we have adopted the mental habit of looking at things from what is assumed to be the searcher's viewpoint, and not (for example) from the document's viewpoint, and hence ultimately from the author's. Perhaps the author has as much of a right to be served as the searcher, i.e., in order that his articles should be retrieved by "relevant readers." And perhaps in that same sense the information store is as interested in searching the searcher as the searcher is in searching the information. After all, the information requirement which activates the searcher to use the system may not be the only one on his mind; it seems to me reasonable that an information user's mind might contain a whole series of standing requests, many of which he is not aware of at the moment of using a retrieval system. A system which functions in an exploratory way, however, has a good chance of fulfilling some of these requests. Cheydleur (3) speaks of "rapport" between the system and the user. The most elemental example of a "system" which permits exploration without barring the mechanics of requesting is permuted title indexing.

Assuming that arguments of the foregoing type can convince people that relevance is not a measurable quantity, like weight or length, by which we can determine the excellence of retrieval to the third decimal place, should we then liquidate relevance as a concept, finding another to put in its place? Probably not. In the first place, people will no more give up the concept of relevance tomorrow morning than they would give up epicycles instantly in response to Copernicus, nor predictability instantly in response to Heisenberg. Relevance is a thought-crutch; with it we may think inaccurately about the retrieval problem, but without it (or something better) we couldn't think at all.

The concept of relevance does more good than ill, unless we take it too literally; it allows us to suboptimize on the machine side of the man-machine interface, which can be a productive thing to do as long as one is not unaware that he is suboptimizing. But in the long run, for the best satisfaction of the searcher, as well as for the satisfaction of all concerned, we must also study the human side of the interface, and study especially both sides in interaction. When the need for exploratory capability in a retrieval system is then acknowledged, what concept will supplant relevance, or at least supplement it? I would suggest "sharpness of separation of the exploratory regions in which the searcher finds documents of interest from those in which he does not find such documents." Note that this criterion gears itself not just to a particular request, nor even to a particular need, but to the full dynamic possibilities of a human using a retrieval system.

"Relevance" will serve its purpose, but will decline as the realization slowly comes that an individual's information need is so complex that it cannot be accurately stated in a simple request. The fact that people do request information in simple terms is a reflection of the inadequacy of both people and systems, and not a reflection of the true structure of the need. The gradually increasing awareness of a human's incapability of stating his true need in simple form will tend to pull the rug out from under many IR system evaluation studies which will have been done in the meanwhile.

References

1. Taylor, Robert S. 1962. The Process of Asking Questions. American Documentation, 13(4): 391-396.

2. Bar-Hillel, Y. 1960. Some Theoretical Aspects of the Mechanization of Literature Searching. Technical Report No. 3. April. U. S. Office of Naval Research, Washington, D. C.

3. Cheydleur, B. F. 1961. Information Retrieval---1966. Datamation, 7(10): 21-25.

4. Stiles, H. E. 1963. Informal communication at a meeting on associative indexing held in January at the National Science Foundation.

Also:

Stiles, H. E. 1961. The Association Factor in Information Retrieval. Journal of the Association for Computing Machinery, 8(2): 271-279.

UNCLASSIFIED

System Development Corporation,
Santa Monica, California
IS RELEVANCE AN ADEQUATE CRITERION IN
RETRIEVAL SYSTEM EVALUATION?
Scientific rept., SP-1262, by L. B. Doyle.
1 July 1963, 6p., 5 refs.
Unclassified report

DESCRIPTORS: Information Retrieval.

Argues that the use of "relevance to search a request" as a criterion of what a system retrieves is, in effect, a suboptimization on the machine side of the man-machine interface, and that the searcher needs an efficient exploratory system rather than a request-implementing system.

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED

UNCLASSIFIED